ADVANCING EARTH AND SPACE SCIENCES

Geophysical Research Letters

RESEARCH LETTER
10.1029/2023GL106434

Phase Neural Operator for Multi-Station Picking of Seismic Arrivals

Hongyu Sun[1], Zachary E. Ross[1], Weiqiang Zhu[1,2], and Kamyar Azizzadenesheli[3]

[1]Seismological Laboratory, California Institute of Technology, Pasadena, CA, USA, [2]Now at Berkeley Seismological Laboratory, University of California, Berkeley, Berkeley, CA, USA, [3]Nvidia Corporation, Santa Clara, CA, USA

Key Points:

- We introduce a multi-station phase picking algorithm, Phase Neural Operator (PhaseNO), that is based on a new machine learning paradigm called Neural Operator
- PhaseNO can use data from any number of stations arranged in any arbitrary geometry to pick phases across the input stations simultaneously
- By leveraging the spatial and temporal contextual information, PhaseNO achieves superior performance over leading baseline algorithms

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

H. Sun,
hongyu-sun@outlook.com

Citation:

Sun, H., Ross, Z. E., Zhu, W., & Azizzadenesheli, K. (2023). Phase neural operator for multi-station picking of seismic arrivals. *Geophysical Research Letters*, *50*, e2023GL106434. https://doi.org/10.1029/2023GL106434

Received 22 SEP 2023
Accepted 26 NOV 2023

Author Contributions:

Conceptualization: Hongyu Sun, Zachary E. Ross
Formal analysis: Hongyu Sun
Funding acquisition: Zachary E. Ross
Investigation: Hongyu Sun
Methodology: Hongyu Sun, Zachary E. Ross, Weiqiang Zhu, Kamyar Azizzadenesheli
Project Administration: Zachary E. Ross
Resources: Zachary E. Ross, Weiqiang Zhu
Software: Hongyu Sun, Weiqiang Zhu
Supervision: Zachary E. Ross

Abstract

Seismic wave arrival time measurements form the basis for numerous downstream applications. State-of-the-art approaches for phase picking use deep neural networks to annotate seismograms at each station independently, yet human experts annotate seismic data by examining the whole network jointly. Here, we introduce a general-purpose network-wide phase picking algorithm based on a recently developed machine learning paradigm called Neural Operator. Our model, called Phase Neural Operator, leverages the spatio-temporal contextual information to pick phases simultaneously for any seismic network geometry. This results in superior performance over leading baseline algorithms by detecting many more earthquakes, picking more phase arrivals, while also greatly improving measurement accuracy. Following similar trends being seen across the domains of artificial intelligence, our approach provides but a glimpse of the potential gains from fully-utilizing the massive seismic data sets being collected worldwide.

Plain Language Summary

Earthquake monitoring often involves measuring arrival times of P- and S-waves of earthquakes from continuous seismic data. With the advancement of artificial intelligence, state-of-the-art phase picking methods use deep neural networks to examine seismic data from each station independently; this is in stark contrast to the way that human experts annotate seismic data, in which waveforms from the whole network containing multiple stations are examined simultaneously. With the performance gains of single-station algorithms approaching saturation, it is clear that meaningful future advances will require algorithms that can naturally examine data for entire networks at once. Here we introduce a multi-station phase picking algorithm based on a recently developed machine learning paradigm called Neural Operator. Our algorithm, called Phase Neural Operator, leverages the spatial-temporal information of earthquake signals from an input seismic network with arbitrary geometry. This results in superior performance over leading baseline algorithms by detecting many more earthquakes, picking many more seismic wave arrivals, yet also greatly improving measurement accuracy.

1. Introduction

Seismic phase detection and picking are fundamental tasks in earthquake seismology, where the aim is to identify earthquakes in the continuous data and measure the arrival times of seismic waves. Historically, human seismic analysts manually labeled earthquake signals and the arrival times of seismic phases by looking for coherent wavefronts on multiple stations and then picking the onset times of P and S waves at each station. Such analysis, however, is subjective, time-consuming, and prone to errors. Considerable effort has been dedicated to developing accurate, automatic, and timely earthquake detection methods, such as short-term average/long-term average (Withers et al., 1998), template matching (Gibbons & Ringdal, 2006; Shelly et al., 2007), and fingerprint and similarity threshold (Yoon et al., 2015). Recent advances in deep learning have greatly improved the accuracy and efficiency of automatic phase picking algorithms (Dokht et al., 2019; Feng et al., 2022; Johnson & Johnson, 2022; Mousavi et al., 2020; Mousavi, Zhu, et al., 2019; Münchmeyer et al., 2022; Perol et al., 2018; Ross et al., 2018; J. Wang et al., 2019; Xiao et al., 2021; Yeck et al., 2021; Zhou et al., 2019; W. Zhu & Beroza, 2018; L. Zhu et al., 2019; W. Zhu, Tai, et al., 2022). However, the single-station detection strategy used in most of the machine-learning detection algorithms can result in failure to detect events with weak amplitude, or mistakenly detect local noise signals with emergence pulses. Indeed, the performance gains of single-station neural phase pickers have rapidly saturated, leading to the question of where the next breakthroughs in phase picking will come from.

Validation: Hongyu Sun
Visualization: Hongyu Sun
Writing – original draft: Hongyu Sun
Writing – review & editing: Hongyu Sun, Zachary E. Ross, Weiqiang Zhu, Kamyar Azizzadenesheli

Across the various domains of artificial intelligence, such as natural language processing and computer vision, the largest gains in performance have come from (a) using ever-larger data sets with increasingly detailed labeling/prediction tasks, (b) making sense of unlabeled data, and (c) incorporating powerful model architectures (e.g., transformers) that are capable of learning to extract information from these very complex data sets. Translating these successes to the phase picking problem would similarly require formulating the problem more generally, in which the goal is to output phase picks only after examining the seismic data for all available sensors in a network. To accomplish such a general formulation, new models are needed that can naturally consider the spatial and temporal context on a variable arrangement of sensors. Although strategies have been proposed to handle the irregular seismic network geometry for earthquake source characterization (van den Ende & Ampuero, 2020; X. Zhang et al., 2022), earthquake early warning (Bloemheuvel et al., 2022; Münchmeyer et al., 2021), and seismic phase association (McBrearty & Beroza, 2023), a generalized network-based phase picker remains an open question.

In this paper, we introduce such an approach for general purpose network-wide earthquake detection and phase picking. Our algorithm, called Phase Neural Operator (PhaseNO), builds on Neural Operators (Kovachki et al., 2023), a recent advance of deep learning models that operate directly on functions rather than finite dimensional vectors. PhaseNO learns infinite dimensional function representations of seismic wavefields across the network, allowing us to accurately measure the arrival times of different phases jointly at multiple stations with arbitrary geometry. We evaluate our approach on real-world seismic data sets and compare its performance with state-of-the-art phase picking methods. We demonstrate that PhaseNO outperforms leading baseline algorithms by detecting many more earthquakes, picking many more phase arrivals, yet also greatly improving measurement accuracy. Overall, our approach demonstrates the power of leveraging both temporal and spatial information for seismic phase picking and improving earthquake monitoring systems.

2. Method: Phase Neural Operator

We introduce an operator learning model for network-wide phase picking (see Text S1 in Supporting Information S1). PhaseNO is designed to learn an operator between infinite-dimensional function spaces on a bounded physical domain. The input function is a seismic wavefield observed at some arbitrary collection of points in space and time, $f(x, y, t)$, and the output function is a probability mask $g(x, y, t)$ that indicates the likelihood of P- and S-wave arrivals at each point (x, y, t). A powerful advantage of Neural Operators over classical Neural Networks is that they are discretization-invariant, meaning that the input and output functions can be discretized on a different (arbitrary) mesh every time a solution is to be evaluated, without having to re-train the model. This critical property allows for Neural Operators to be evaluated at any point within the input physical domain, enabling phase picking on a dynamic seismic network with different geometries.

We combine two types of Neural Operators to naturally handle the mathematical structure of seismic network data. For the temporal information, we use Fourier Neural Operator (FNO) layers (Li et al., 2020a), which are ideal for cases in which the domain is sure to be discretized on a regular mesh, because fast Fourier transforms are used to quickly compute a solution. Since seismograms are mostly sampled regularly in time, FNO can efficiently process and encode seismograms. For the spatial information, our sensors are generally not on a regular mesh, and so we instead use Graph Neural Operators (GNO, Li et al., 2020b) to model the relationship of seismic waveforms at different stations. This type of neural operator is naturally able to work with irregular sensors, as it uses message passing (Gilmer et al., 2017) to aggregate features from multiple stations and construct an operator with kernel integration.

Figure 1 summarizes the PhaseNO architecture. The model is composed of multiple blocks of operator layers in which FNO and GNO are sequentially connected and repeated several times, allowing for sufficient communications and exchange of spatiotemporal information between all stations in a seismic network. Skip connections are used to connect the blocks, resulting in a U-shape architecture. The skip connection directly concatenates FNO results on the left part of the model with GNO results on the right without going through deep layers, which improves convergence and allows for deeper, more overparameterized models.

3. Results

3.1. Performance Evaluation

In this study, we benchmark the performance of PhaseNO against three leading baseline models (see Text S3 in Supporting Information S1): EQTransformer (Mousavi et al., 2020), PhaseNet (W. Zhu & Beroza, 2018), and

10.1029/2023GL106434

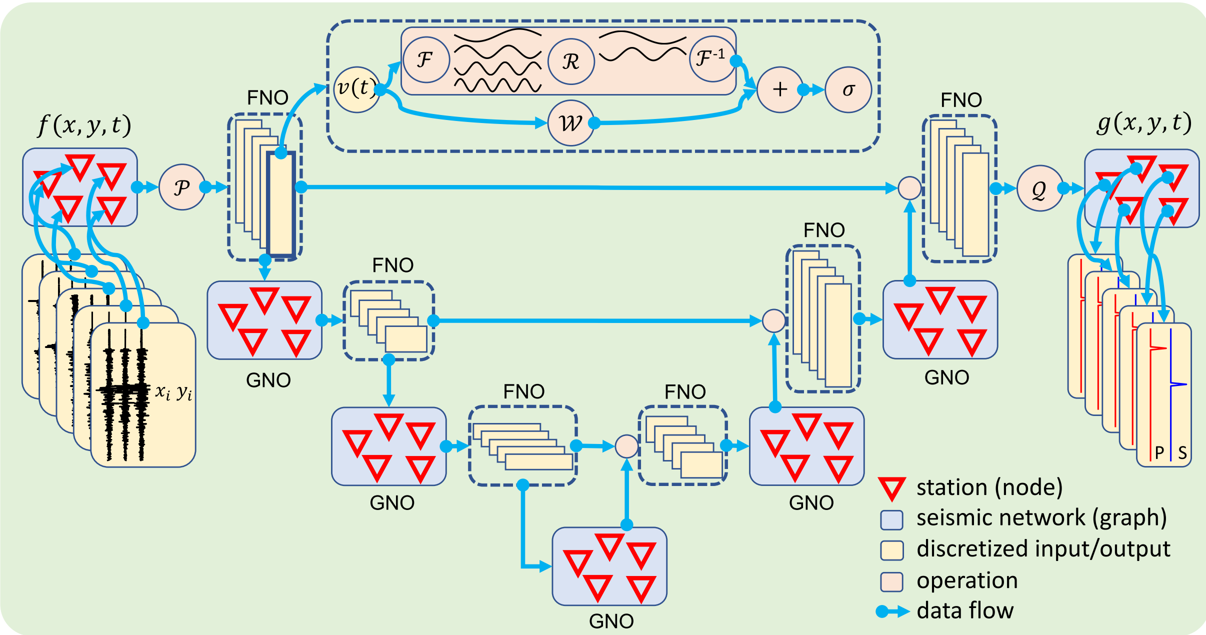

Figure 1. *Phase Neural Operator architecture*. The model consists of multiple Fourier Neural Operator (FNO) and Graph Neural Operators (GNO) layers that are sequentially connected and repeated. $\mathcal{P}$ and $\mathcal{Q}$ are up- and down-projections parameterized by neural networks. The model uses seismograms from a seismic network containing multiple stations with an arbitrary geometry as the input and predicts the probabilities of P-phase and S-phase arrival times for all input stations. Station locations are encoded as two channels of the input, in addition to three channels carrying the three-component waveforms. The relative locations (x_i, y_i) between stations can be used to learn weights as edge features in a graph (see Text S2 in Supporting Information S1).

EdgePhase (Feng et al., 2022). We trained PhaseNO on an earthquake data set from the Northern California Earthquake Data Center (NCEDC) spanning the period 1984–2019 (see Text S4 in Supporting Information S1), that is, the same training data set as PhaseNet. We evaluated PhaseNO and each baseline model on an out of sample test data set for the period 2020 containing 43,700 P/S picks of 5,769 events. We choose the time window for each sample based on their pre-trained models: 30 s for PhaseNO and PhaseNet, and 60 s for EQTransformer and EdgePhase. Positions of picks are randomly placed in the middle 30 s of the time window. For all of the models, P- and S-picks were determined from peaks in the predicted probability distributions by setting a pre-determined threshold. Each model used a distinct threshold as the one maximizing the F1 score to ensure the models compared under their best conditions (Figure 2a; Figure S1 in Supporting Information S1).

Our method results in the highest F1 scores for both P- and S-waves, being 0.99 and 0.98 respectively. This is in addition to having the highest optimal thresholds (0.70 for P and 0.65 for S) of all the models tested (Table S1 in Supporting Information S1). Given that similar labeling strategies were used for training the baselines (Gaussian for PhaseNet and triangular for the other models), a higher threshold indicates that PhaseNO has a higher confidence level for detecting and picking seismic arrivals than other methods. When true picks are unavailable to determine the optimal threshold for a particular test data set based on F1 scores (i.e., the trade-off between correct and false phases), PhaseNO is able to minimize false detection and give more picks compared with other methods with the same pre-determined threshold. The two single station picking models, PhaseNet and EQTransformer, have similar F1 scores, but the former has higher recall and the latter has higher precision. EdgePhase is built on EQTransformer and has better performance in terms of the precision-recall curves. However, the phase picks are less precise in terms of time residuals (Table S1 and Figure S2 in Supporting Information S1). PhaseNO detects more true positives, fewer false negatives, and fewer false positive picks than the other deep-learning models at almost all signal-to-noise ratio (SNR) levels (Figure S3 in Supporting Information S1). Despite generating more picks, PhaseNO results in the smallest mean absolute error for both P and S phases. Overall, PhaseNO achieves the best performance on all six metrics, with one minor exception. The standard deviation of P phase residuals for PhaseNO is 0.01 s (one time step) larger than PhaseNet. It should be noted that the newly detected phases by PhaseNO are likely to be more challenging cases as their signal-to-noise levels are lower, and thus result in slightly increased standard deviation.

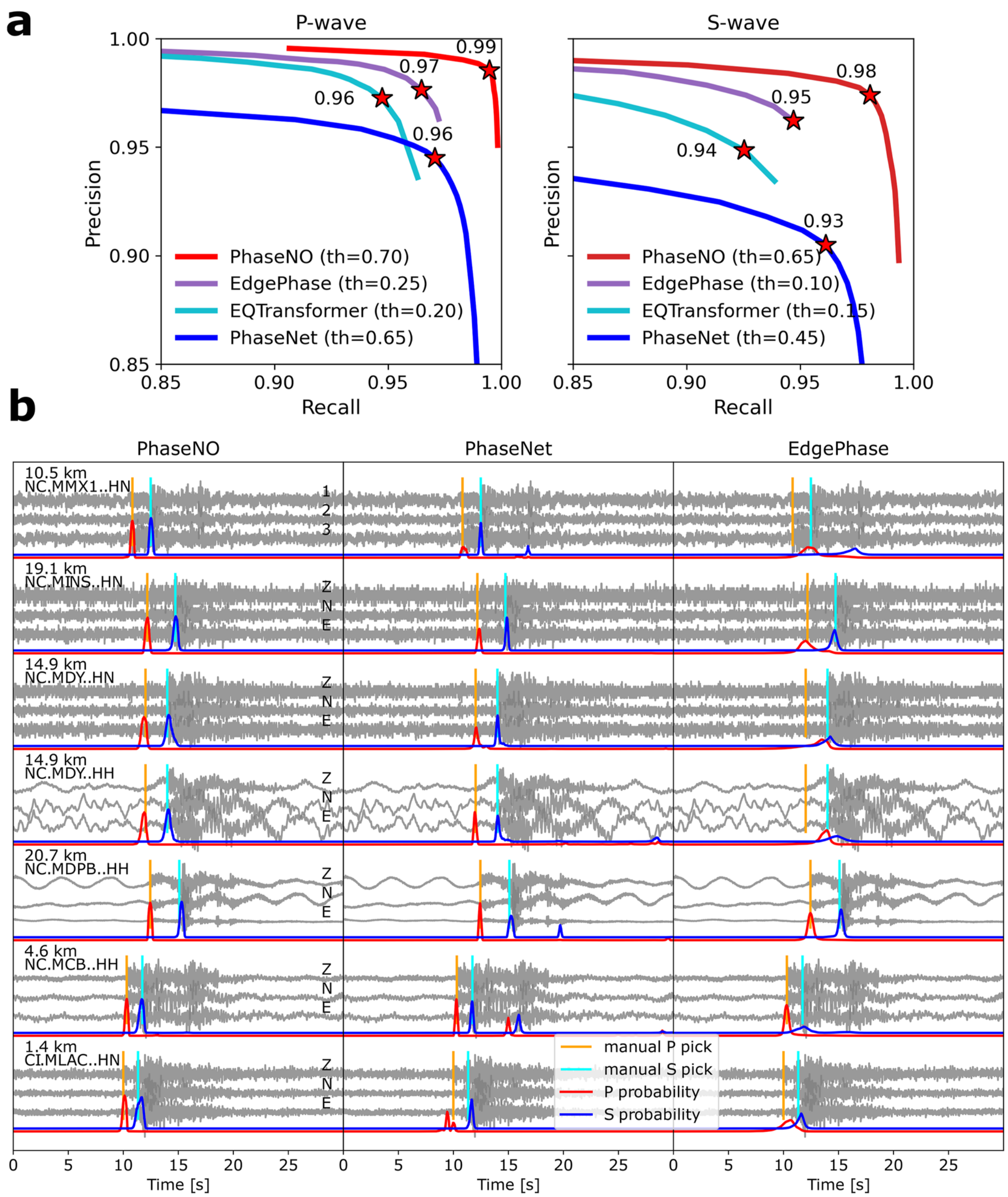

Figure 2. *Performance evaluation on the NCEDC2020 test data set*. (a) Precision-recall curves. The best threshold (th) for each model on this test data set is selected based on the maximum F1 scores (stars labeled on the curves) that models achieve (Figure S1 in Supporting Information S1). (b) Event nc71112909 with a magnitude of 0.43. The station name and epicentral distance are shown on waveforms.

We compare the predicted probability distributions of each neural phase picker for several representative events (Figure 2b, Figures S4 and S5 in Supporting Information S1). PhaseNO works very well on different event magnitudes, instrument types, and waveform shapes. PhaseNet generates some false positive picks that are removed by multi-station methods (PhaseNO and EdgePhase); however, EdgePhase also generates many false negatives. Through exchanging temporal and spatial information multiple times, PhaseNO effectively prevents false picks while improving the detection ability of true picks. PhaseNO successfully finds picks on low SNR waveforms by leveraging contextual information from other stations.

S-phases generally exist in the coda of P-phases and are more challenging to find. Thus, more labeling errors from human analysts are expected on S phases than P phases. For instance, in Figure S4a in Supporting Information S1, three of the models generate consistent S picks, but the predicted peaks are systematically offset from the manual picks on this event. For these example cases, PhaseNO shows significant improvement in S-phase picking and generates higher probabilities than the other methods. Moreover, the width of the picks predicted by PhaseNO may represent the degree of difficulty in picking the phases from the waveforms, even though the same label width is used in baselines and our method. Picks with high probabilities may have wider distribution if the waveforms have low SNR. Also, our model can handle the waveforms with more than one pick existing in a sample (Figure S5 in Supporting Information S1).

3.2. Application to the 2019 Ridgecrest Earthquake Sequence

We tested the detection performance and generalization ability of PhaseNO on the 2019 Ridgecrest earthquake sequence. We downloaded continuous waveform data for EH, HH, and HN sensors for the period 4 July 2019 (15:00:00) to 10 July 2019 (00:00:00) at 20 Southern California Seismic Network (SCSN) stations, which is a total of 36 distinct sensors. Each of these sensors is treated as a distinct node in the graph, even if they are co-located (Table S2 in Supporting Information S1). Waveform data are divided into hourly streams with a sampling rate of 100 Hz. This is a challenging data set due to the overlap of numerous events. Since no ground-truth catalog is available for the continuous data, we evaluated our results by comparing them with catalogs produced by SCSN, PhaseNet, and two template matching studies (Ross et al., 2019; Shelly, 2020).

We first divided the entire seismic network into two parts and constructed two graphs for every hour of data, due to the increased computational cost with the number of nodes in a graph (Figure S6 in Supporting Information S1). The 36 nodes were randomly divided into two graphs with 18 nodes. Continuous data were cut into a 30-s time window with an overlap of 10 s, resulting in 180 predictions for 1-hr data on 18 nodes. After preprocessing, PhaseNO predicted the probabilities of earthquake phases on 18 nodes at once. We compare representative waveforms with probabilities predicted by PhaseNO and PhaseNet (Figures 3 and 4, and Figures S7–S11 in Supporting Information S1). Both models show great generalization ability, as these waveforms were recorded outside of the training region. Our model works very well on continuous data, especially when there is more than one event in a 30-s time window, when the event is located at any position of the window, and when the waveform has different shapes with low SNR. Owing to the learned waveform consistency among multiple stations, PhaseNO detects much more picks with meaningful moveout patterns than PhaseNet.

After prediction, we determined phase picks using a threshold of 0.3 for both P and S phases. PhaseNO detected 693,266 P and 686,629 S arrival times, while PhaseNet found 542,793 P and 572,991 S arrival times with the same threshold and the same stations. We evaluated the accuracy of the detected picks by comparing the arrival times with manually reviewed picks from SCSN (Figure S12 in Supporting Information S1). The standard deviation of the pick residuals between SCSN and PhaseNO was 0.10 s for P phases and 0.14 s for S phases, calculated from 118,746 P picks and 96,247 S picks. The standard deviation, however, was slightly higher than those with PhaseNet (0.08 s for P from 106,061 picks and 0.13 s for S from 88,438 picks). Since the newly detected picks are more challenging cases with low fidelity, it is reasonable for PhaseNO to show a larger travel time difference.

We convert candidate phase detections into events by phase association with GaMMA (W. Zhu, McBrearty, et al., 2022). We set a minimum of 17 picks per event to filter out low-quality associations. This results in PhaseNet detecting 21,748 events with 37.54 picks per event, whereas PhaseNO detects 26,176 events with 39.37 picks per event (Figure 5a). Many of the unassociated picks are probably a consequence of our strict filtering criteria during association, rather than false detections. With the same association hyperparameters, the additional 4,428 events highlight the advancement of PhaseNO for earthquake detection. Despite the increased number of events, PhaseNO shows high detection quality with around two more picks per event compared to PhaseNet, even though they are smaller events in general (Figure S13 in Supporting Information S1). GaMMA calculates magnitudes for events detected by PhaseNO and PhaseNet, and they both show linear Gutenberg-Richter distributions (Figure 5b). Indeed, our results have fewer microearthquakes than the template matching catalog by Ross et al. (2019). Since microearthquakes usually have limited propagation ranges and can only be recorded by several stations, they would have been filtered out during association and thus not shown on the frequency-magnitude distribution. Moreover, event locations determined by GaMMA are generally consistent between PhaseNO and PhaseNet catalogs (Figure 5c), confirming that the additional events by PhaseNO are reasonable detections of real earthquakes.

Furthermore, we treat the manually reviewed SCSN catalog as a baseline and evaluate how many earthquakes were successfully recovered. We consider that two events are matched if they occur within 3 s from each other.

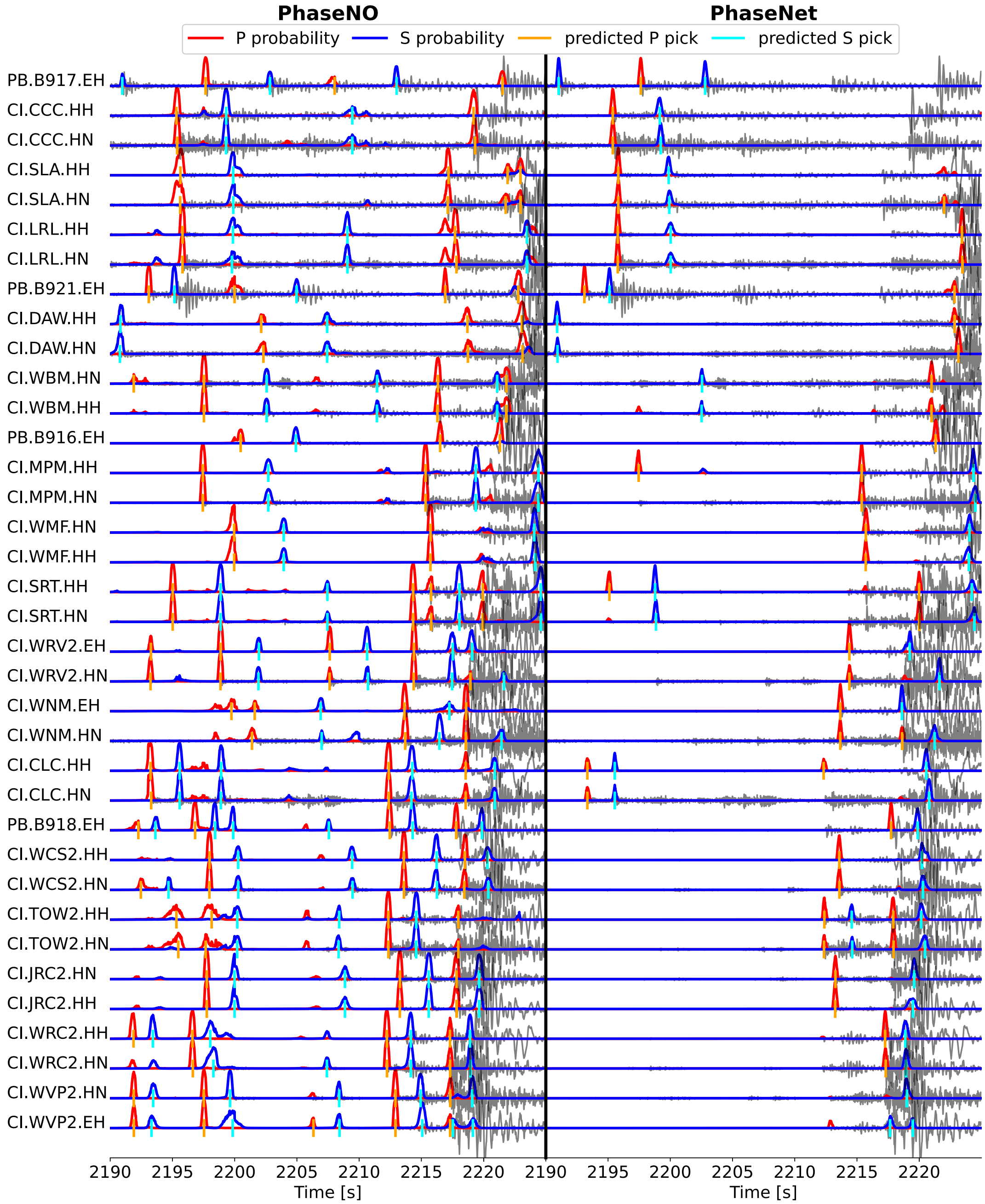

Figure 3. Example results for a 35-s window during the 2019 Ridgecrest earthquake sequence.

With such criteria, Shelly, Ross et al., and PhaseNet matched around 81%, 86%, and 88% events, respectively. In comparison, with our strict filtering criteria during association, PhaseNO catalog totaling 26,176 events matched approximately 94% events in the SCSN catalog (10,673 of 11,389) with additional events, indicating the highest recall score of PhaseNO. PhaseNO consistently detects more events than PhaseNet, SCSN, and Shelly's template matching catalog (Shelly, 2020) over time and approaches the number of earthquakes reported by another more detailed template matching catalog (Ross et al., 2019). Moreover, PhaseNO achieves a much more stable detection with the greatest number of events found when the M_w 7.1 mainshock occurred (Figure 5a) and with the gradually reduced seismicity rate afterward, indicating the power of the method to illuminate complex earthquake sequences. Examples of events and associated picks detected by PhaseNO can be found in Supporting Information S1 (Figures S14–S16 in Supporting Information S1).

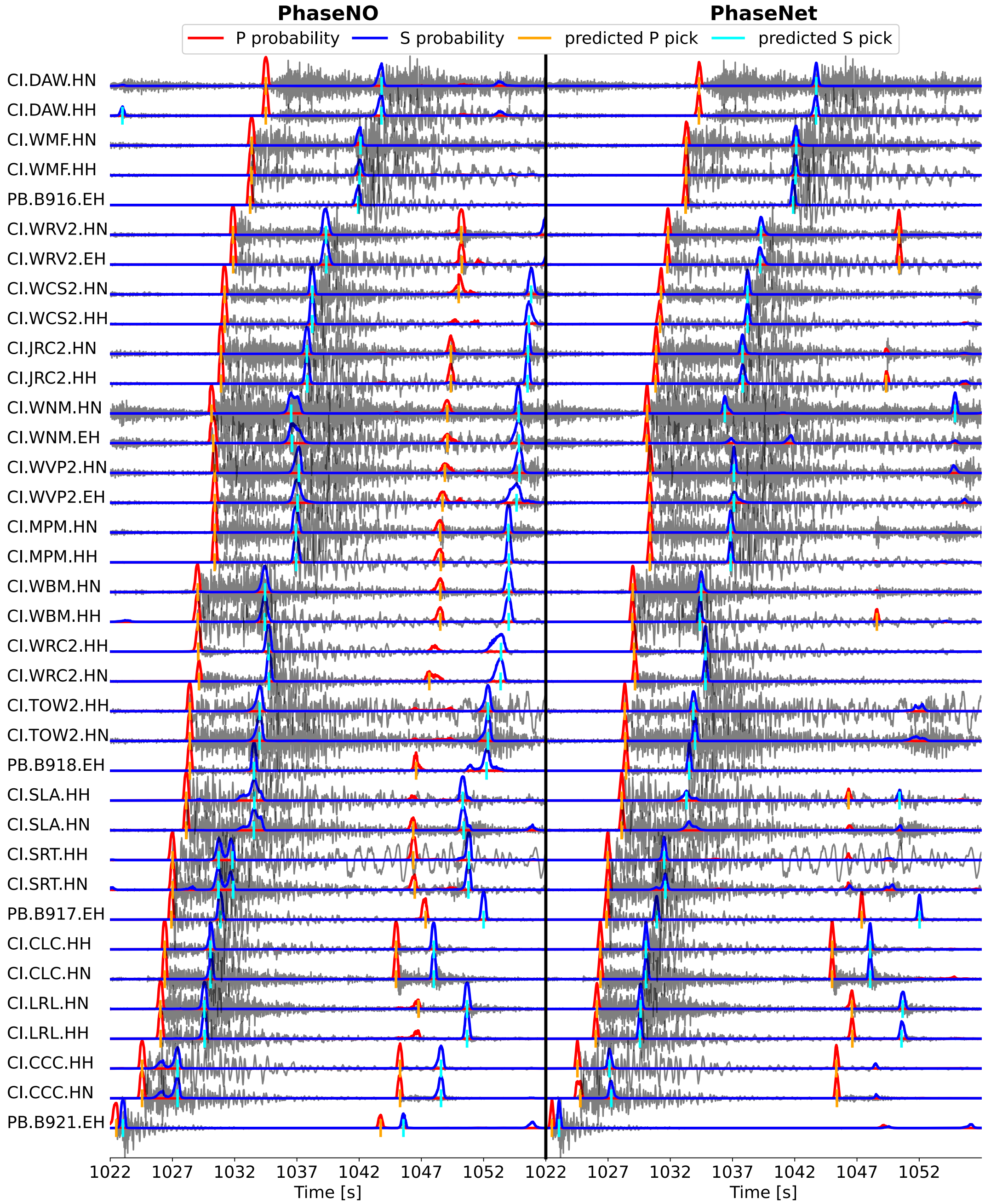

Figure 4. Example results for a short window during the 2019 Ridgecrest earthquake sequence. More examples can be found in Supporting Information S1 (Figures S7–S11 in Supporting Information S1).

It should be noted that our catalog differs from those of the SCSN and template matching catalogs in the number of stations and association algorithms. However, picks from PhaseNO and PhaseNet are detected on the exact same stations and then associated with GaMMA, providing the fairest comparison. Two post-processing hyperparameters, the threshold in phase picking and the minimum number of picks associated with an event, control the total number of earthquakes in a catalog. A lower threshold and a smaller association minimum provide more events, despite likely more false positive events (Table S3 in Supporting Information S1). PhaseNO consistently detects more events than PhaseNet using the same hyperparameters, pointing out the importance of leveraging the spatial information in addition to the temporary information for phase picking.

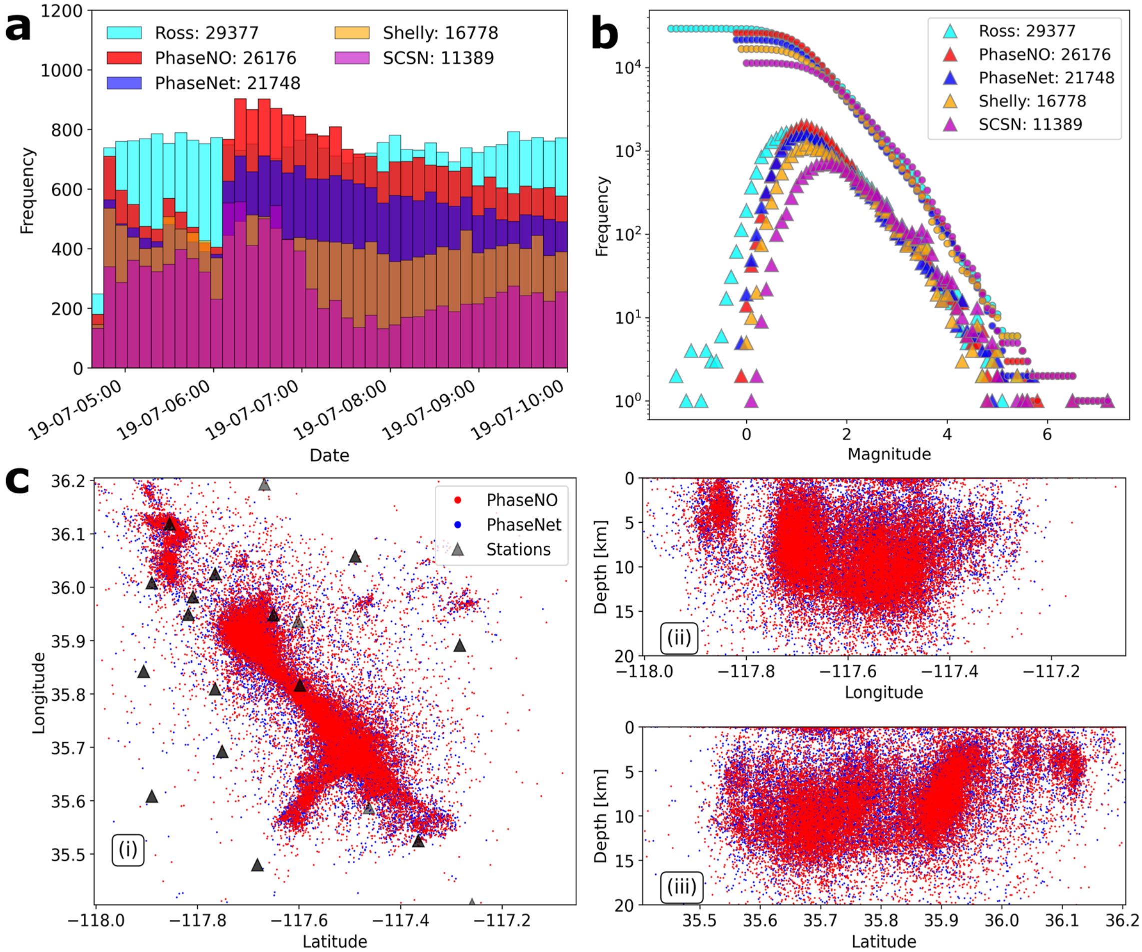

Figure 5. *Comparison of earthquake catalogs of the 2019 Ridgecrest earthquake sequence*. (a) Earthquake number. (b) Frequency-magnitude distributions. (c) Earthquake hypocenters.

4. Discussion and Conclusions

With a fixed model architecture, PhaseNO can handle seismic networks with arbitrary geometries; we demonstrated this by training on the Northern California Seismic Network and evaluated the model on the Southern California Seismic Network, without retraining. This is a critical property of the Neural Operator class of models, which can learn in infinite dimensions.

PhaseNO shows several distinctive characteristics in terms of network design. Compared to most of the currently popular detection algorithms (deep-learning or traditional methods), PhaseNO mimics human learning and decision making by using context from the whole seismic network, rather than seismograms at a single station. By consulting information and searching for consistent waveforms from surrounding stations, PhaseNO greatly improves phase picking on low SNR data, especially S phases that usually are hidden in the coda of P phases.

Apart from the characteristics in the spatial domain, PhaseNO has a unique ability to identify phases from temporal information. The well-known transformer architecture that has brought about major successes in natural language processing (Vaswani et al., 2017) can be viewed as a special case of Neural Operators (Kovachki et al., 2023). Just as EQTransformer uses an attention mechanism to investigate global dependencies, PhaseNO supervises the global features with kernel integrals in space and time. Like PhaseNet, PhaseNO adopts a U-shape architecture with skip connections, which improves model convergence and allows for a deeper model design with greater expressiveness.

Compared to EdgePhase, a multi-station picking model, our model uses multiple GNO layers, a type of Neural Operator that allows for kernel integration over the network to extract rich spatial features. Each GNO layer is inserted between two FNO layers, forcing the exchange of information between spatial and temporal domains. We also encode station locations as node features to weight the message constructed between nodes. Additionally, instead of building a graph based on geographic distances and only selecting neighboring nodes within a certain distance from the target node, we construct a graph using all nodes in a seismic network. All these modifications contribute to maximizing the usage of spatial features for phase picking.

A major limitation to PhaseNO, however, is the dependence of memory usage on the number of stations in one prediction. Spatial information is exchanged between all pairs of nodes in a graph; therefore, the computational cost scales quadratically with the number of nodes, with complexity $O(n^2)$. Hence, we suggest selecting a subset of stations from the entire large seismic network for one prediction until all stations have been processed before moving to the next time segment of continuous data, like the procedure described in the Ridgecrest example. If the seismic network covers a wide range of areas, we may select stations based on *k*-means clustering (Lloyd, 1982). In this way, we can greatly accelerate the prediction procedure and save memory usage, particularly when there are many stations and when the computational resources are limited.

Data Availability Statement

Version v1.0.0 of PhaseNO and the pre-trained model are preserved at Sun (2023). The training and test data are from Northern California Earthquake Data Center (NCEDC, 2014). The data of the 2019 Ridgecrest earthquake sequence can be accessed from Southern California Earthquake Data Center (SCEDC, 2013), and Plate Boundary Observatory Borehole Seismic Network (NCEDC, 2014).

Acknowledgments

We thank the Editor Daoyuan Sun, Christopher W. Johnson and an anonymous reviewer for constructive comments on the manuscript. ZER is grateful to the David and Lucile Packard Foundation for supporting this study through a Packard Fellowship.

References

Bloemheuvel, S., van den Hoogen, J., Jozinović, D., Michelini, A., & Atzmueller, M. (2022). Graph neural networks for multivariate time series regression with application to seismic data. *International Journal of Data Science and Analytics*, *16*, 317–332. https://doi.org/10.1007/s41060-022-00349-6

Dokht, R. M. H., Kao, H., Ryan, V., & Smith, B. (2019). Seismic event and phase detection using time-frequency representation and convolutional neural networks. *Seismological Research Letters*, *90*(2A), 481–490. ISSN 0895-0695, 1938-2057. https://doi.org/10.1785/0220180308

Feng, T., Mohanna, S., & Meng, L. (2022). EdgePhase: A deep learning model for multi-station seismic phase picking. *Geochemistry, Geophysics, Geosystems*, *23*(11), e2022GC010453. ISSN 1525-2027. https://doi.org/10.1029/2022GC010453

Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, *165*(1), 149–166. ISSN 0956540X, 1365246X. https://doi.org/10.1111/j.1365-246X.2006.02865.x

Gilmer, J., Schoenholz, S. S., Riley, P. F., Oriol Vinyals, & Dahl, G. E. (2017). Neural message passing for quantum chemistry. Retrieved from http://arxiv.org/abs/1704.01212

Johnson, C. W., & Johnson, P. A. (2022). EQDetect: Earthquake phase arrivals and first motion polarity applying deep learning. https://doi.org/10.1002/essoar.10511191.1

Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2023). Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, *24*(89), 1–97. Retrieved from http://jmlr.org/papers/v24/21-1524.html

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020a). Fourier neural operator for parametric partial differential equations. https://doi.org/10.48550/ARXIV.2010.08895

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020b). Neural operator: Graph kernel network for partial differential equations. Retrieved from http://arxiv.org/abs/2003.03485

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/tit.1982.1056489

McBrearty, I. W., & Beroza, G. C. (2023). Earthquake phase association with graph neural networks. *Bulletin of the Seismological Society of America*, *113*(2), 524–547. https://doi.org/10.1785/0120220182

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer - An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, *11*(1), 3952. ISSN 2041-1723. https://doi.org/10.1038/s41467-020-17591-w

Mousavi, S. M., Zhu, W., Sheng, Y., & Beroza, G. C. (2019). CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection. *Scientific Reports*, *9*(1), 10267. ISSN 2045-2322. https://doi.org/10.1038/s41598-019-45748-1

Münchmeyer, J., Bindi, D., Leser, U., & Tilmann, F. (2021). The transformer earthquake alerting model: A new versatile approach to earthquake early warning. *Geophysical Journal International*, *225*(1), 646–656. https://doi.org/10.1093/gji/ggaa609

Münchmeyer, J., Jack, W., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., et al. (2022). Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. *Journal of Geophysical Research: Solid Earth*, *127*(1), e2021JB023499. ISSN 2169-9313, 2169-9356. https://doi.org/10.1029/2021JB023499

NCEDC. (2014). Northern California Earthquake Data Center. UC Berkeley Seismological Laboratory [Dataset]. https://doi.org/10.7932/NCEDC

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2), e1700578. ISSN 2375-2548. https://doi.org/10.1126/sciadv.1700578

Ross, Z. E., Idini, B., Jia, Z., Stephenson, O. L., Zhong, M., Wang, X., et al. (2019). Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence. *Science*, *366*(6463), 346–351. ISSN 0036-8075, 1095-9203. https://doi.org/10.1126/science.aaz0109

Ross, Z. E., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, *108*(5A), 2894–2901. ISSN 0037-1106, 1943-3573. https://doi.org/10.1785/0120180080

SCEDC. (2013). Southern California Earthquake Data Center. Caltech [Dataset]. https://doi.org/10.7909/C3WD3xH1

Shelly, D. R. (2020). A high-resolution seismic catalog for the initial 2019 Ridgecrest earthquake sequence: Foreshocks, aftershocks, and faulting complexity. *Seismological Research Letters*, *91*(4), 1971–1978. ISSN 0895-0695, 1938-2057. https://doi.org/10.1785/0220190309

Shelly, D. R., Beroza, G. C., & Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, *446*(7133), 305–307. ISSN 0028-0836, 1476-4687. https://doi.org/10.1038/nature05666

Sun, H. (2023). PhaseNO (v1.0.0) [Software]. Zenodo. https://doi.org/10.5281/zenodo.10224301

van den Ende, M. P. A., & Ampuero, J.-P. (2020). Automated seismic source characterization using deep graph neural networks. *Geophysical Research Letters*, *47*(17), e2020GL088690. https://doi.org/10.1029/2020gl088690

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. Retrieved from http://arxiv.org/abs/1706.03762

Wang, J., Xiao, Z., Chang, L., Zhao, D., & Yao, Z. (2019). Deep learning for picking seismic arrival times. *Journal of Geophysical Research: Solid Earth*, *124*(7), 6612–6624. ISSN 2169-9313, 2169-9356. https://doi.org/10.1029/2019JB017536

Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., & Trujillo, J. (1998). A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bulletin of the Seismological Society of America*, *88*(1), 95–106. ISSN 1943-3573, 0037-1106. https://doi.org/10.1785/BSSA0880010095

Xiao, Z., Wang, J., Chang, L., Li, J., Zhao, L., & Yao, Z. (2021). Siamese earthquake transformer: A pair-input deep-learning model for earthquake detection and phase picking on a seismic array. *Journal of Geophysical Research: Solid Earth*, *126*(5), e2020JB021444. ISSN 2169-9313, 2169-9356. https://doi.org/10.1029/2020JB021444

Yeck, W. L., Patton, J. M., Ross, Z. E., Hayes, G. P., Guy, M. R., Ambruz, N. B., et al. (2021). Leveraging deep learning in global 24/7 real-time earthquake monitoring at the National Earthquake Information Center. *Seismological Research Letters*, *92*(1), 469–480. ISSN 0895-0695, 1938-2057. https://doi.org/10.1785/0220200178

Yoon, C. E., O'Reilly, O., Bergen, K. J., & Beroza, G. C. (2015). Earthquake detection through computationally efficient similarity search. *Science Advances*, *1*(11), e1501057. ISSN 2375-2548. https://doi.org/10.1126/sciadv.1501057

Zhang, X., Reichard-Flynn, W., Zhang, M., Hirn, M., & Lin, Y. (2022). Spatiotemporal graph convolutional networks for earthquake source characterization. *Journal of Geophysical Research: Solid Earth*, *127*(11), e2022JB024401. https://doi.org/10.1029/2022jb024401

Zhou, Y., Han, Y., Kong, Q., & Zhou, S. (2019). Hybrid event detection and phase-picking algorithm using convolutional and recurrent neural networks. *Seismological Research Letters*, *90*(3), 1079–1087. ISSN 0895-0695, 1938-2057. https://doi.org/10.1785/0220180319

Zhu, L., Peng, Z., McClellan, J., Li, C., Yao, D., Li, Z., & Fang, L. (2019). Deep learning for seismic phase detection and picking in the aftershock zone of 2008 M7.9 Wenchuan Earthquake. *Physics of the Earth and Planetary Interiors*, *293*, 106261. ISSN 00319201. https://doi.org/10.1016/j.pepi.2019.05.004

Zhu, W., & Beroza, G. C. (2018). PhaseNet: A deep-neural-network-based seismic arrival time picking method. *Geophysical Journal International*, *216*(1), 261–273. ISSN 0956-540X, 1365-246X. https://doi.org/10.1093/gji/ggy423

Zhu, W., McBrearty, I. W., Mousavi, S. M., Ellsworth, W. L., & Beroza, G. C. (2022). Earthquake phase association using a Bayesian Gaussian mixture model. *Journal of Geophysical Research: Solid Earth*, *127*(5). ISSN 2169-9313, 2169-9356. https://doi.org/10.1029/2021JB023249

Zhu, W., Tai, K. S., Mousavi, S. M., Bailis, P., & Beroza, G. C. (2022). An end-to-end earthquake detection method for joint phase picking and association using deep learning. *Journal of Geophysical Research: Solid Earth*, *127*(3), e2021JB023283. ISSN 2169-9313, 2169-9356. https://doi.org/10.1029/2021JB023283

References From the Supporting Information

Hendrycks, D., & Gimpel, K. (2020). Gaussian error linear units (GELUs). Retrieved from http://arxiv.org/abs/1606.08415

Mousavi, S. M., Sheng, Y., Zhu, W., & Beroza, G. C. (2019). STanford EArthquake Dataset (STEAD): A global data set of seismic signals for AI. *IEEE Access*, *7*, 179464–179476. ISSN 2169-3536. https://doi.org/10.1109/ACCESS.2019.2947848

Rahman, M. A., Ross, Z. E., & Azizzadenesheli, K. (2022). U-no: U-shaped neural operators. arXiv preprint arXiv:2204.11127.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2019). Dynamic graph CNN for learning on point clouds. *ACM Transactions on Graphics*, *38*(5), 1–12. ISSN 0730-0301, 1557-7368. https://doi.org/10.1145/3326362

Zhang, M., Liu, M., Tian, F., Wang, R., & Zhu, W. (2022). LOC-FLOW: An end-to-end machine learning-based high-precision earthquake location workflow. *Seismological Society of America*, *93*(5), 2426–2438. https://doi.org/10.1785/0220220019